



Complete nucleotide sequence and genome organization of Calibrachoa mottle virus (CbMV)—A new species in the genus *Carmovirus* of the family *Tombusviridae*[☆]

Anju Gulati-Sakhuja, Hsing-Yeh Liu^{*}

USDA-ARS 1636 E Alisal Street, Salinas, CA 93905, USA

ARTICLE INFO

Article history:

Received 12 August 2009

Received in revised form 4 November 2009

Accepted 8 November 2009

Available online 13 November 2009

Keywords:

CbMV

Calibrachoa mottle virus

Carmovirus

Genome sequence

Phylogenetic analysis

Tombusviridae

ABSTRACT

Complete genomic sequence of the viral RNA of Calibrachoa mottle virus (CbMV) has been determined. The CbMV genome has a positive-sense single-stranded RNA of 3919 nucleotides in length and encodes five open reading frames (ORFs). ORF1 encodes a protein with predicted molecular weight of 28 kDa (p28). ORF2 extends through the amber stop codon of ORF1 to give a protein with a predicted molecular weight of 87 kDa (p87). The readthrough domain of p87 contains the GDD motif common to RNA-dependent RNA polymerases (RdRp). ORF3 and ORF4 encode two small overlapping polypeptides of 8 kDa (p8) and 9 kDa (p9), respectively. The 3'-proximal ORF5 encodes a capsid protein (CP) of 37 kDa (p37). The untranslated 5'- and 3'-terminal regions are composed of 34 and 234 non-coding nucleotides, respectively. Comparisons of amino acid sequences of the ORFs of CbMV with members of *Tombusviridae* show that CbMV is closely related to members of the genus *Carmovirus*. Phylogenetic analyses based on the amino acid sequences of RdRp and coat protein and nucleotide sequences of the whole genome reveal that CbMV forms a subgroup with several carmoviruses. Therefore, the genome organization, physico-chemical properties, sequence alignments and phylogenetic analysis support the classification of CbMV as a new species in the genus *Carmovirus*, family *Tombusviridae*.

Published by Elsevier B.V.

1. Introduction

A new virus infecting *Calibrachoa*, an important new horticultural plant in Europe and in the United States, was first reported and isolated by Liu et al. (2003). The infected plants showed interveinal chlorosis and mottling or blotching on the leaves of susceptible cultivars at stressed conditions (Liu et al., 2003; White, 2003). The virus was transmitted mechanically to *Chenopodium amaranticolor*, *C. capitatum*, *C. quinoa*, *Nicotiana benthamiana*, and *N. clevelandii* plants, but not transmitted by green peach aphid (*Myzus persicae*), sweet potato whitefly (*Bemisia tabaci*), silverleaf whitefly (*B. argentifolii*), greenhouse whitefly (*Trialeurodes vaporariorum*), or banded-wing whitefly (*T. abutilonea*) (Liu et al., 2003). Purified virions from *Calibrachoa* plants were spherical particles of 29–31 nm in diameter with a single-stranded (ss) RNA of approximately 4.0 kb. The capsid protein of approximately 41 kDa was resolved by SDS-PAGE. The double stranded (ds)RNA profile consistently revealed

one major band of about 4.0 kb and three minor bands of approximately 3.1, 1.6 and 1.3 kb. Using primers designed based on the *Pelargonium flower break virus* (PFBV) RNA-dependent RNA polymerase (RdRp) sequence (EMBL L34289) (Morozov et al., 1995), a small RT-PCR product of about 300 bp was amplified from viral RNA and sequenced. A database search using BLAST algorithm (Altschul et al., 1990) showed homology with highly conserved motif of RdRp associated with *Carmovirus* and shared 94% identity with *Carnation mottle virus* (CarMV, X02986). However, *Calibrachoa* virus and CarMV were distinct serologically and based on respective host ranges. Based on its natural host symptoms and available sequence data, the virus was named Calibrachoa mottle virus (CbMV) and was classified as a member of the genus *Carmovirus* in the family *Tombusviridae* (Liu et al., 2003).

The genus name, *Carmovirus*, is derived from its first sequenced member CarMV. So far, the complete nucleotide (nt) sequences for 15 carmovirus species are available in GenBank that includes: *Angelonia flower break virus* (AnFBV) (Adkins et al., 2006), *Cardamine chlorotic fleck virus* (CCFV) (Skotnicki et al., 1993), CarMV (Guilley et al., 1985), *Cowpea mottle virus* (CPMoV) (You et al., 1995), *Galinsoga mosaic virus* (GaMV) (Ciuffreda et al., 1998), *Hibiscus chlorotic ringspot virus* (HCRSV) (Huang et al., 2000), *Japanese iris necrotic ring virus* (JINRV) (Takemoto et al., 2000), *Melon necrotic spot virus* (MNSV) (Riviere and Rochon, 1990), *Nootka lupine vein-clearing*

[☆] The mention of firm names or trade products in this article does not imply endorsement or recommendation by the United States Department of Agriculture over other firms or similar products not mentioned.

^{*} Corresponding author. Tel.: +1 831 755 2826; fax: +1 831 755 2814.

E-mail address: hsingyeh.liu@ars.usda.gov (H.-Y. Liu).

virus (NLVVCV) (Robertson et al., 2007), *Pea stem necrosis virus* (PSNV) (Suzuki et al., 2002), *Pelargonium flower break virus* (PFBV) (Rico and Hernandez, 2004), *Pelargonium line pattern virus* (PLPV) (Castano and Hernandez, 2005), *Saguaro cactus virus* (SCV) (Weng and Xiong, 1997), *Turnip crinkle virus* (TCV) (Carrington et al., 1989) and *Soybean yellow mottle mosaic virus* (SYMMV) (Nam et al., 2008). TCV is the only carmovirus for which refined structural and *in vitro* assembly studies have been performed.

The genome organization of the carmoviruses is quite compact and contains five open reading frames (ORFs) with most of them overlapping with each other. Mutagenesis studies in TCV have demonstrated that all five ORFs are essential for infectivity (Hacker et al., 1992). ORF1 terminates with an amber stop codon and encodes a protein of 26–29 kDa. Readthrough of the amber codon results in the synthesis of a protein of 83–89 kDa (ORF2). Both proteins are putative subunits of the viral replicase (Hacker et al., 1992). The 3′ proximal gene encodes viral coat protein (CP) which varies from 37–42 kDa for different viruses. Two small centrally located ORFs, a characteristic of carmoviruses, encode proteins that have been shown in TCV to be involved with viral cell-to-cell movement (movement proteins or MPs) (Hacker et al., 1992). *In vitro* translation studies of CarMV (Carrington and Morris, 1985, 1986) showed that the first two ORFs were expressed from the viral genomic RNA. Similar studies in SCV and TCV has also shown that CP and MPs are synthesized from 1.5 kb and 1.7 kb subgenomic RNAs, respectively. The 5′ non-coding region varies from 6 nt in PLPV to 134 nt in PSNV with no extensive sequence homology. There is also no evidence for the presence of cap at 5′ end of carmovirus genomes. The 3′ non-coding region of carmoviruses varies from 224 to 406 nt and lacks either a poly(A) tract or a terminal tRNA-like structure.

To confirm the identity of CbMV as a member of the genus carmovirus, and determine its relationship with other carmoviruses, cDNA clones representing the complete genome of CbMV were prepared and sequenced. Comparisons of its whole nucleotide sequence and deduced amino acid sequences of major ORFs of CbMV with selected members of *Tombusviridae* have also been discussed.

2. Materials and methods

2.1. Virus propagation, purification and dsRNA extraction

The CbMV was originally isolated from symptomatic leaves of *Calibrachoa* plants and maintained in *N. benthamiana* by mechanical inoculations (Liu et al., 2003). Virions were extracted and purified from infected *N. benthamiana* plants as described by Liu et al. (2003). Viral-associated dsRNA was isolated according to Valverde et al. (1990) using 7 g of infected *Calibrachoa* plant tissues.

2.2. cDNA cloning and nucleotide sequencing

First-strand of cDNA was synthesized from purified CbMV viral RNA, using random hexamers and SuperScript III Reverse Transcriptase (Invitrogen, Carlsbad, CA), per manufacturer's instructions, followed by second-strand synthesis using the Riboclone cDNA synthesis kit (Promega, Madison, WI) with some modifications. The reaction mixture was incubated at 14 °C for 2 h followed by addition of 10 units T4 DNA Polymerase and incubation at 16 °C for 5 min. The reaction was stopped by adding 10 μl of 0.5 M EDTA. After phenol:chloroform extraction and ethanol precipitation, double stranded cDNA was directly cloned into the vector pCR-Blunt II-TOPO using Zero Blunt TOPO PCR cloning kit (Invitrogen, Carlsbad, CA). Ligation products were transformed and amplified in Mach-I competent cells (Invitrogen) as per manufacturer's protocol. The plasmid DNAs from selected clones

were extracted with Wizard Plus SV Miniprep kit (Promega) and sequenced bi-directionally by TACGen (Richmond, CA) using M13 forward and M13 reverse primers. The resulting nucleotide and derived amino acid sequences of different clones were compared using the BLASTN and BLASTP (Altschul et al., 1990, 1997) and assembled by using Mac Vector 7.0 AssemblyLIGN (Oxford Accelrys Inc., San Diego, CA) programs.

To fill the gaps between sequences, PCR primers were designed at different regions of the assembled sequence to obtain overlapping sequence. PCR was performed using Platinum Pfx DNA Polymerase (Invitrogen) as per manufacturer's protocol using standard amplification profiles consisted of initial denaturation at 94 °C for 5 min, followed by 35 cycles of 94 °C for 30 s, 50–60 °C for 45 s, and 68 °C for 1–5 min; ending with 10 min at 68 °C. The amplified RT-PCR products were cloned in pCR-Blunt II-TOPO vector and sequenced.

For recalcitrant sections of CbMV, dsRNA with added reverse primer was denatured with 20 mM methyl mercury for 6 min at room temperature, followed immediately by heating at 65 °C for 4 additional minutes and incubation on ice for 1 min. Reverse transcription was performed using Superscript III. PCR was performed with primers designed on the assembled sequence using PrimeSTAR HS DNA polymerase (Takara, Madison, WI) as per manufacturer's protocol. The amplified products were cloned into pCR-Blunt II-TOPO vector for sequencing.

To obtain clones representing the 3′ end of viral RNA, dsRNA was polyadenylated using *Escherichia coli* poly(A) polymerase (Ambion, Foster city, CA) following the manufacturer's instructions. After phenol:chloroform extraction and ethanol precipitation, the polyadenylated RNA was denatured with 20 mM methyl mercury and reverse transcribed with oligo-dT (5′-GGCCACGCTCGACTAGTAC(T)₁₇-3′) of 3′ Rapid Amplification of cDNA Ends (RACE) System (Invitrogen). The RT product was PCR amplified with PrimeSTAR HS DNA polymerase (Takara) using oligo-dT primer and specific primer 5′-CCACAAAAGAGGATAATCAGACCAC-3′, which is homologous to nt 3702–3727 in the complete CbMV sequence. The PCR product was cloned and subsequently sequenced. The 3′-end sequence was verified by adding poly(C) tail (using poly(A) polymerase and rCTP) and 3′ RACE system using oligo-dG (5′-TACTACTCATATTAATATA(G)₁₄-3′) and the same forward (nt 3702–3727) CbMV primers. Similarly, for amplification of the 5′-terminal region, poly(A) and poly(C) tailed dsRNAs were utilized for RT-PCR amplification with oligo-dT or oligo-dG reverse primers, respectively, and two individual oligonucleotide forward primers (5′-CTCGTCATCCTTTCTTCTTCTTC-3′, complementary to nt 339–316 and 5′-CACTTCAGCTTTAGCATACTTGGG-3′, complementary to nt 243–220) derived from near the 3′-end of the complete CbMV sequence of the minus strand as described by Tzanetakis and Martin (2004).

2.3. Sequence and phylogenetic analysis

Sequence data was assembled and analyzed using the Mac Vector 7.0 AssemblyLIGN program. Identification of open reading frames (ORFs) and protein translation were conducted with MacVector 7.0 ORF analysis programs. The theoretical isoelectric points (pI) of ORFs 3 and 4 were computed using Expasy Proteomics tools (<http://www.expasy.org/>). Multiple sequence alignments and pairwise comparisons of the CbMV sequence with selected members of *Tombusviridae* were done with ClustalW (v1.4) using alignment parameters with a gap opening penalty of 10, a gap extension penalty of 0.2, and the BLOSUM protein weight matrix. All nucleotide and amino acid sequences of other viruses were obtained from GenBank database and the details are provided in Table 1. Phylogenetic trees were created using the UPGMA method,

Table 1
Percentage of nucleotide and amino acid sequence identity of Calibrachoa mottle virus (accession #GQ244431) to carmoviruses and selected members of other genera in the family Tombusviridae^{a,b}.

Virus ^b	(accession #)	Nucleotide identity	Amino acid identity					
			P28	P87	PRTD ^c	P8	P9	CP
<i>Carmovirus</i>								
AnFBV	(DQ219415)	52	35	49	56	26	35	29
CCFV	(L16015)	49	25	40	47	30	29	31
CarMV	(AF192772)	51	30	49	59	30	47	33
CPMoV	(U20976)	47	19	38	47	27	31	25
GaMV	(Y13463)	47	19	36	43	28	16	26
HCRSV	(X86448)	49	25	42	50	24	33	29
JINRV	(D86123)	48	23	41	49	22	25	28
MNSV	(M29671)	45	19	39	50	28	22	24
NLVCV	(EF207438)	52	44	52	56	34	26	33
PFBV	(AJ514833)	52	45	54	59	23	26	39
PLPV	(AY613852)	50	27	42	49	27	15	28
PSNV	(AB086951)	48	19	41	50	28	16	23
SCV	(U72332)	54	39	52	59	23	36	43
SYMMV	(FJ457015)	48	22	39	47	23	30	28
TCV	(M22445)	49	25	41	48	20	–	30
Members of other genera in <i>Tombusviridae</i>								
<i>Aureusvirus</i>								
PoLV	(AJ243370)	44	14	30	39	– ^d	–	23
<i>Avenavirus</i>								
OCSV	(X83964)	45	18	29	35	16	–	21
<i>Dianthovirus</i>								
CRSV	(L18870)	44	19	26	32	–	–	25
RCNMV	(J04357)	45	23	28	35	–	–	23
SCNMV	(L07884)	44	17	28	34	–	–	24
<i>Machlomovirus</i>								
MCMV	(X14736)	43	18	31	45	15	–	20
<i>Necrovirus</i>								
BBSV	(AF452884)	45	17	31	37	20	21	16
LWSV	(X94560)	44	16	32	30	23	19	19
OLV1	(X85989)	47	20	37	45	29	20	16
TNV-A	(M33002)	46	22	37	44	20	16	19
TNV-D	(U62546)	46	18	32	38	18	27	19
<i>Panicovirus</i>								
PMV	(U55002)	45	15	30	42	23	–	17
<i>Tombusvirus</i>								
AMCV	(X62493)	41	17	28	39	–	–	21
CIRV	(X85215)	42	18	29	38	–	–	22
CNV	(M25270)	42	20	29	38	–	–	20
GALV	(AY830918)	41	18	28	38	–	–	23
TBSV	(M21958)	41	18	28	38	–	–	20

^a Homologies with less than 25% identity over the length of the protein are not considered related.

^b Sequences used for comparison were retrieved from Genbank sequence database. The viral abbreviations are: AnFBV, *Angelonia flower break virus*; AMCV, *Artichoke mottled crinkle virus*; BBSV, *Beet black scorch virus*; CCFV, *Cardamine chlorotic fleck virus*; CIRV, *Carnation Italian ringspot virus*; CarMV, *Carnation mottle virus*; CNV, *Cucumber necrosis virus*; CPMoV, *Cowpea mottle virus*; CRSV, *Carnation ringspot virus*; GALV, *Grapevine Algerian latent virus*; GaMV, *Galinsoga mosaic virus*; HCRSV, *Hibiscus chlorotic ringspot virus*; JINRV, *Japanese iris necrotic ring virus*; LWSV, *Leek white stripe virus*; MCMV, *Maize chlorotic mottle virus*; MNSV, *Melon necrotic spot virus*; NLVCV, *Nootka lupine vein-clearing virus*; OCSV, *Oat chlorotic stunt virus*; OLV1, *Olive latent virus 1*; PFBV, *Pelargonium flower break virus*; PLPV, *Pelargonium line pattern virus*; PMV, *Panicum mosaic virus*; PolV, *Pothos latent virus*; PSNV, *Pea stem necrosis virus*; RCNMV, *Red clover necrotic mosaic virus*; SCV, *Saguaro cactus virus*; SCNMV, *Sweet clover necrotic mosaic virus*; SYMMV, *Soybean yellow mottle mosaic virus*; TBSV, *Tomato bushy stunt virus*; TCV, *Turnip crinkle virus*; TNV-A, *Tobacco necrosis virus A*; TNV-D, *Tobacco necrosis virus D*.

^c PRTD: post-readthrough domain (p87 domain following the ORF1 stop codon).

^d Homologue is absent.

Systematic ties, Poisson-Correction distance for RdRP and CP amino acid sequences and Tamura-Nei for complete nucleotide sequence with Gap sites ignored. Bootstrap analysis consisted of 10,000 replications. Prediction of RNA secondary structure of 3' non-translated regions of the virus was performed using Mfold software (Zuker, 2003) at default settings with flat exterior loop type.

3. Results

3.1. Genome organization of CbMV

The complete genomic RNA of CbMV has been sequenced and deposited in Genbank with accession number GQ244431. The total

length of RNA is 3919 nt with a GC content of 46%. The genomic organization of CbMV has been shown in Fig. 1. The CbMV genome contains five adjacent or overlapping open reading frames (ORFs). The 5'-proximal ORF starts at nt 37 and proceeds up to nt 777. It encodes a protein of 28 kDa (246 amino acids, p28) that terminates with an *amber* codon at nt 775–777 surrounded by a sequence AAAUAGGGG. The suppression of the p28 *amber* codon would result into ORF2, which potentially encodes an in-frame readthrough protein of 87 kDa (762 amino acids, p87 replicase) and terminates with an *ochre* codon at nt 2323–2325. The p87 of CbMV contains amino acid sequence Gly-Asp-Asp, commonly referred to as the GDD motif of RNA-dependent RNA polymerase, at nt 1710–1718.

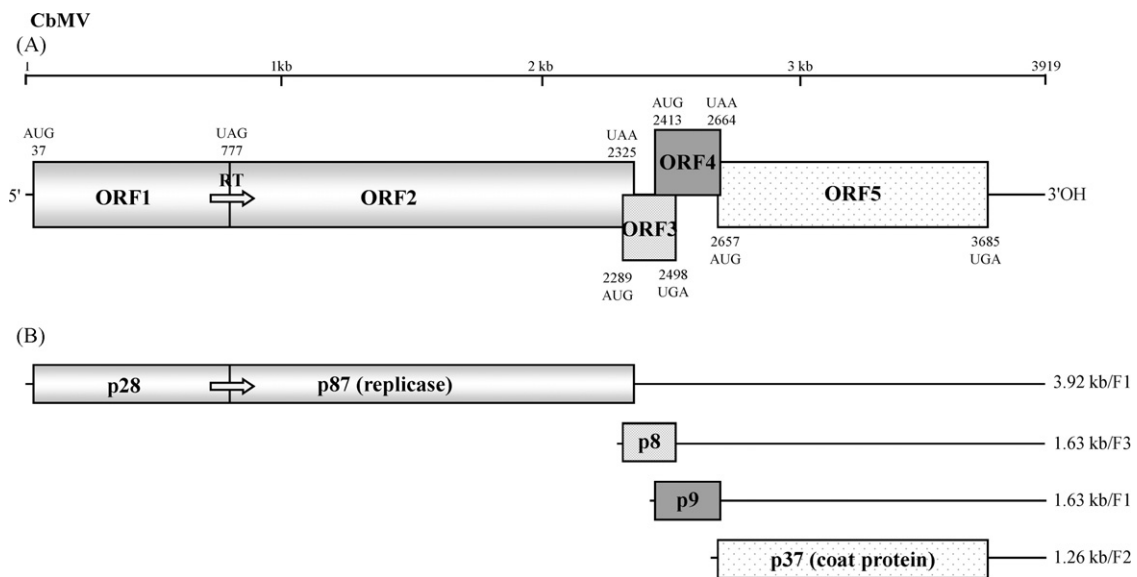


Fig. 1. Genome organization (A) and predictable translation strategy (B) of Calibrachoa mottle virus (CbMV). Boxes with solid line represent the ORFs with the size of the respective proteins (or readthrough product) indicated within. Overlapping boxes represent overlapping ORFs. ORF1 is in-frame with ORF2 and terminated with an amber stop codon, UAG. The numbers above and below the boxes represent the nucleotide position of the start and stop codons defining each ORF. Arrow depicts a readthrough (RT) of amber termination codon UAG of the first ORF to a large fusion protein. The approximate sizes of genomic RNA (3.92 kb) and two subgenomic RNAs (1.63 and 1.26 kb) that serve as templates for translation are indicated on the right (B). F1–F3 indicate the positive-sense reading frame from which proteins are translated.

Downstream of ORF2 are two centrally located small overlapping ORFs, ORF3 and ORF4 (nt 2289–2498 and nt 2413–2664, respectively). The initiation codon of ORF3 is present at position 2289–2291, overlapping 37 nt of ORF2 and is present in a different reading frame from ORF1/2. It encodes a small polypeptide of 69 amino acids with a molecular mass of 8 kDa (p8) and terminates with an *opal* stop codon at 2496–2498 nt. Amino acid sequence analysis of p8 shows a putative RNA-binding domain (KDKNKLSGKLTAAKAVANE) present between its 25 and 43 amino acids (nt 2360–2417). ORF4 is present in the same frame as the replicase and encodes another small protein of 9 kDa (83 amino acids, p9), beginning 86 nt upstream of the stop codon of ORF3. This terminates with an *ochre* stop codon at nt 2662–2664. Both p8 and p9 are highly basic proteins with a pI of 9.92 and 9.10, respectively.

The 3'-proximal ORF encoding a 342 amino acids polypeptide with molecular mass of 37 kDa (p37, coat protein (CP)) starts at nt 2657 and terminates at nt 3685 with an *opal* terminator codon. This ORF5 overlaps ORF4 by 8 nt and is present in a different frame than the other ORFs. The CbMV genome contains 36 non-coding nt at the 5'-end and 234 non-coding nucleotides at non-polyadenylated 3'-end. The identity of both the ends was confirmed by poly(rA) and poly(rC) tailing of CbMV RNA. The sequences from three different cDNA clones (for each reaction) corresponding to this region were identical and confirms the sequence of 3'-end.

In addition to five ORFs described above, a small ORF of 51 amino acids is also predicted from the sequence. It starts at nt 2962 and ends at nt 3117, encoding a protein of 6 kDa. This ORF is nested within ORF5 and is in the same frame as replicase and p9.

3.2. Comparison of CbMV genome with related viruses

The nucleotide and derived amino acid sequences of CbMV were compared with 15 other distinct virus species in the genus *Carmovirus* where complete genomic sequences were available (Table 1). The CbMV complete nucleotide sequence exhibits identity percentages ranging from 45 to 54% with the sequences of other carmoviruses. The highest overall nucleotide identities are with SCV (54%) followed by AnFBV, NLVCV and PFBV (52%). When comparing overall percent amino acid identity for each gene, the

replicase (p87) is the most conserved gene ranging from 54% (PFBV) to 36% (GaMV) and p8 is the least conserved, ranging from 34% (NLVCV) to 20% (TCV). The maximum homology is in the post-readthrough domain (PRTD) of the putative RNA-dependent RNA polymerase (RdRp) gene, ranging from 59% in CarMV, NLVCV and SCV to 43% in GaMV, compared to the pre-readthrough domain of replicase (p28) that shares only 45% (PFBV) to 19% (CPMoV, GaMV, MNSV and PSNV) identity. This shows that PRTD of RdRp is more conserved than p28. The other two genes, p9 and p37 (CP) show intermediate amino acid sequence identity, p9 ranges from 47% in CarMV to 15% in PLPV, and CP ranges from 43% in SCV to 23% in PSNV. Therefore, coat protein of CbMV is less conserved than RdRp (Table 1). CbMV CP is only 43%, 39%, and 33% identical to those of SCV, PFBV, and NLVCV where as p87 is 54% and 52% identical in PFBV and SCV and NLVCV, respectively. No extensive homology is observed within 5' non-coding region of carmoviruses and other members of *Tombusviridae* (data not shown). The 3'-terminal sequence of CbMV RNA may fold into a stem-loop consisted of about 34 nt (5'-UUAUCAGGGGACUGUUGAGGAGUCUCCCCGCCG-3').

The complete nucleotide sequence and derived amino acid sequences of the various CbMV ORFs are compared with other members of *Tombusviridae* (Table 1). The CbMV whole genome nucleotide sequences shows 47% (OLV1) to 41% (AMCV, GALV and TBSV) identities with those of selected members of other genus of *Tombusviridae* (Table 1). The amino acid sequence of the CbMV replicase gene (p87) also exhibits 37% to 26% identities with those of other viruses (Table 1). The PRTD exhibits the highest percent identities (45–30%) as compared to 23–14% in the pre-readthrough domain (p28). No significant identity is found with p8 and p9. CbMV CP sequence homology with several other viruses of the family is much lower than p87. Although, several members of family had significant similarity with CbMV in the p87 region, not in CbMV CP and CP from other viruses (Table 1).

3.3. Phylogenetic analysis

This analysis focused on the generation of phylogenetic trees based on multiple alignments of complete genome sequences and deduced amino acid sequences of different genes (RdRp and CP) of

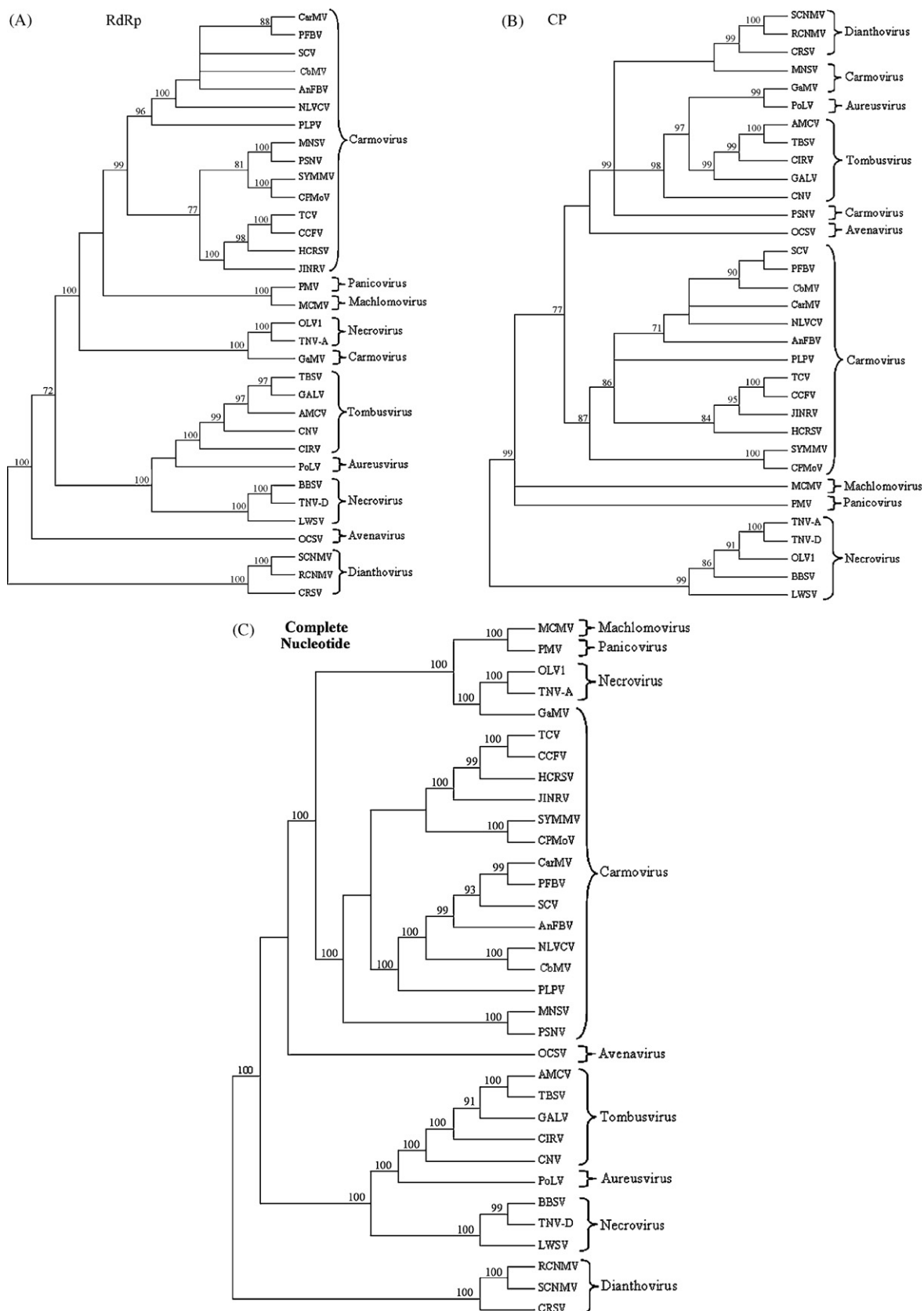


Fig. 2. Phylogenetic analysis of Calibrachoa mottle virus (accession #GQ244431) and selected members of the family Tombusviridae based on the amino acid sequences of RNA-dependent RNA polymerase (A) and coat protein (B) and nucleotide sequences of the whole genome (C). Sequences were aligned using CLUSTALW (v1.4) with gap opening and extension penalties of 10.0 and 0.2, respectively and BLOSUM series protein weight matrix for amino acid sequences. Phylogenetic trees were constructed using the UPGMA method with 10,000 Bootstrap replicates. The numbers at the nodes indicate the percentage at which the grouping occurred; only values >50% are shown. Virus name abbreviations and accession numbers of the sequences used to construct the tree are the same as in Table 1.

CbMV and 32 members of family *Tombusviridae* listed in Table 1. In all cases, CbMV is grouped with CarMV, PFBV, SCV and AnFBV in the clade containing carmoviruses (Fig. 2A–C). Inspection of p28 revealed the similar pattern (data not shown). The alignment of proteins homologous to CbMV p8 from different carmoviruses and other members of the family *Tombusviridae* (Table 1) were used to construct a phylogenetic tree. The corresponding tree showed no clear taxonomic grouping of CbMV with carmoviruses or other members within the family (data not shown).

4. Discussion

Viruses in genus *Carmovirus* have icosahedral virions of about 30–35 nm in diameter. The genomic RNA consists of a linear, monopartite, positive-sense ssRNA of 3803–4266 nucleotides. Carmoviruses have narrow natural host range and are transmitted mechanically but not by aphids or whiteflies. CbMV resembles carmoviruses in particle size and morphology, dsRNA profile, host ranges, and properties of particles in sap (Liu et al., 2003).

Advances in biochemical and molecular biology techniques have changed the system of plant virus classification. Earlier, the viruses were classified based on their physico-chemical and biological properties. More recently, many viral genomes have been sequenced. Genome similarities and phylogenetic tree analysis based on single genes, together with vector similarities and serological properties has facilitated the classification of plant viruses into families and genera by the Plant Virus Subcommittee of the International Committee on Taxonomy of Viruses (ICTV). The present research on the genome organization, physico-chemical properties, sequence alignments and phylogenetic analysis further confirmed the classification of CbMV in the genus *Carmovirus*, family *Tombusviridae*.

The genome organization of CbMV, with replicase, followed by two small overlapping ORFs and the coat protein at the 3' end of the genome is similar to reported carmoviruses. There are, however, a few exceptions: in CarMV, JINRV and PFBV, double readthrough events that extend the translation of polymerase gene into ORF3 to produce p98 or p99 are reported (Guilley et al., 1985; Rico and Hernandez, 2004; Takemoto et al., 2000), and in MNSV, ORF3 and ORF4 are connected by an in-frame amber codon that could result in the expression of either two separate proteins (p7a and p7b), or as readthrough of ORF3 (p14) (Riviere and Rochon, 1990). Readthrough protein p14 in MNSV was not found to be functional *in vivo* (Genoves et al., 2006). No such features are observed in CbMV. Additional ORFs of unknown function have also been postulated in some carmoviruses, and confirmed in HCRSV (Diez et al., 1999; Huang et al., 2000; Rico and Hernandez, 2004; You et al., 1995). In CbMV, an ORF (p6), nested within CP gene is predicted. Database searches of CbMV p6 did not show any similarity with any known protein of other viruses. The existence and biological significance of this protein, if produced in infected plants, has yet to be established.

Comparisons of amino acid sequences of CbMV with selected members of *Tombusviridae* showed extensive identities between the proteins encoded by ORF2, the readthrough domain of replication associated protein. The p87 of CbMV resembles those of the carmoviruses, and are most closely to PFBV, NLVCV and SCV. Both the products of this most 5'-proximal ORF and its readthrough product have been shown to be involved with expression and/or regulation of the RdRp domain in viruses (Hacker et al., 1992; Skuzeski et al., 1990; White et al., 1995). Significant identities of the post-readthrough proteins suggest that they may have a common evolutionary ancestry (Koonin and Dolja, 1993; Weng and Xiong, 1997). The lack of sequence homology between the pre-readthrough domains of replicase suggest that the two domains evolved at a different rate, or independently and recombined later (Riviere and Rochon, 1990; Russo et al., 1994).

Among carmoviruses and members of other genus in *Tombusviridae*, readthrough of the amber stop codon to produce a putative RdRp is a characteristic feature of their genome organization (Harrell et al., 2002; Morris and Carrington, 1988; Robertson et al., 2007). The sequence surrounding the amber stop codon is consistent with the consensus sequence, AA(A/G)UAGC(G/U)(G/A) required for efficient readthrough (Harrell et al., 2002; Robertson et al., 2007; Skuzeski et al. 1990). The most highly conserved GDD motif surrounded by hydrophobic amino acids in CbMV p87 is similar to all RNA-dependent RNA polymerases of RNA viruses (Argos, 1988; Kamer and Argos, 1984). Site directed mutagenesis in TMV has shown that the expression of the putative viral RdRp is regulated by the downstream cistron from leaky amber codon (Ishikawa et al., 1991). The product of ORF1 could not be assigned a specific role in viral replication (Habibi and Symons, 1989).

The CbMV p8 and p9 proteins shared limited homology with its respective counterparts present in other carmoviruses. The involvement of these proteins with virus cell-to-cell movement in TCV and CarMV (Carrington et al., 1989; Guilley et al., 1985) has been confirmed by *in vitro* mutagenesis studies in TCV (Hacker et al., 1992). The putative RNA-binding domain present in the CbMV p8 protein is similar to the one identified in CarMV (Marcos et al., 1999) and NLVCV (Robertson et al., 2007). No such obvious nucleic acid binding sequences are observed in CbMV p9. Further, CbMV p9 is a hydrophobic protein predicted to contain transmembrane domains that are absent in p8. Similar results have been reported in CarMV p9 with helical transmembrane domains that would target p9 to membranes and cell wall (Marcos et al., 1999). The variation in movement protein of carmoviruses may reflect the diverse host ranges of each carmovirus (Weng and Xiong, 1997).

The stem-loop structure at 3' end of CbMV is similar to one reported in CarMV (Canizares et al., 2001), HCRSV (Huang et al., 2000), PFBV (Rico and Hernandez, 2004) and PLPV (Castano and Hernandez, 2005). Analysis of 3' terminal sequence comparison of genomic RNAs of AFBV, CarMV, CCFV, CPMoV, MNSV, SCV, NLVCV, PFBV, and TCV also showed that similar structures could be formed. Such 3' stem-loop structure has been shown to be involved with *in vivo* replication of genomic- and subgenomic-RNAs of virus (Bringle et al., 1999; Carpenter and Simon, 1998; Castano and Hernandez, 2005; Fabian et al., 2003; Fabian and White, 2004; Song and Simon, 1995; Stupina and Simon, 1997).

The current classification of viruses is based on single gene analysis of the replicase or coat protein gene. Recently, a comprehensive whole genome phylogeny of family *Tombusviridae*, produced by Stuart et al. (2004), showed that carmoviruses did not form an exclusive and coherent group. Instead, they were subgrouped with other genera within the family, suggesting a weak association of genes within the genus *Carmovirus*. Therefore, the comparison of whole genomes was recommended to provide a correct analysis and relationship. The whole genome analysis using nucleotide sequences of CbMV and other viruses exhibited clustering of CbMV with CarMV, PFBV, SCV, AnFBV, NLVCV and PLPV within carmoviruses (Fig. 2C). Similar phylogenetic relationship was observed with RdRp and CP trees of CbMV (Fig. 2A and B).

The physico-chemical properties of CbMV, e.g. spherical particles with single-stranded RNA of approximately 4.0 kb and a single coat protein of approximately 41 kDa, resembles other carmoviruses (Liu et al., 2003). Both these properties are confirmed by CbMV complete genome sequence analysis where CbMV contains 3919 nucleotides and an ORF encoding coat protein of ~37 kDa.

In a previous study, gel analysis of dsRNA from CbMV infected *Calibrachoa* plants revealed the presence of four dsRNA species of about 4.0, 3.1, 1.6 and 1.3 kb in size (Liu et al., 2003). Although, these dsRNAs were not identified individually by sequencing, it is inferred that the largest (4.0 kb) and two smaller dsRNAs (1.6 and 1.3 kb) represent the whole genome and subgenomic RNAs,

respectively. The theoretical predicted sizes of dsRNAs from current sequenced genome of CbMV are approximately 3.92, 1.63, and 1.26 kb (Fig. 1). However, these speculative values need to be confirmed by appropriate replication studies. The dsRNA of approximately 3.1 kb noticed during gel analysis was smaller than genomic RNA but larger than subgenomic RNAs. The presence of such additional dsRNAs is rare but not uncommon for carmoviruses. Two minor dsRNAs of approximately 2.9 and 3.3 kb were reported in addition to a genomic RNA (4.0 kb) and two subgenomic RNAs (1.5 and 1.7 kb) in PFBV (Diez et al., 1999). Similarly, four major bands of sizes 4.0, 1.75, 1.60 and 1.31 kb were observed on AnFBV dsRNA gel analysis (Adkins et al., 2006). Northern blot analysis of AnFBV viral RNA also revealed the presence of 3.96, 2.59, 1.30, 1.10 and 0.93 kb RNAs (Adkins et al., 2006). Additional studies are required to determine the role of 3.1 kb dsRNA in CbMV gene expression strategy.

Like other carmoviruses, CbMV was shown to be serologically distinct from CarMV (Liu et al., 2003). The homology in the CP amino acid sequences has been shown to be related with serological diversity among viral species (Koenig and Gibbs, 1986; Koenig et al., 1988; Li et al., 1993). In CbMV, the amino acid sequence of CP gene revealed low percentage identity with other viral species (Table 1). According to the VIIIth Report of the International Committee on Taxonomy of Viruses (ICTV) (Lommel et al., 2005), the species demarcation criteria for carmoviruses is <41% and <52% identities of the CP and replicase amino acid sequences, respectively. Pairwise amino acid sequence comparison of CbMV CP gene with that of CarMV exhibited only 33% identity. Similar comparisons of CbMV CP gene with those of other carmoviruses also showed 43% (SCV) to 23% (PSNV) homology (Table 1). High identity between CbMV and SCV shows that the two viruses are most closely related with each other. The replicase gene also demonstrated the relationship of CbMV with SCV, PFBV and NLVCV. However, based on species demarcation criteria percentage for CP and replicase, CbMV can be classified as a distinct species in genus carmovirus.

In conclusion, the complete nucleotide sequence of CbMV has been determined. Based on genome organization, CbMV is comparable to all other carmoviruses. Amino acid sequence comparisons (% identity) of the replicase gene, two small movement genes and the CP gene indicate that CbMV is closely related with carmoviruses. Finally, phylogenetic analysis conducted with different viral proteins and whole genome nucleotides confirms its inclusion within the genus *Carmovirus*, family *Tombusviridae*.

Acknowledgements

The authors thank Laura L. Hladky, Dr. Gail Wisler and Dr. Ruhui Li for their critical review of the manuscript.

References

- Adkins, S., Hammond, J., Gera, A., Maroon-Lango, C.J., Sobolev, I., Harness, A., Zeidan, M., Spiegel, S., 2006. Biological and molecular characterization of a new carmovirus isolated from Angelonia. *Phytopathology* 96, 460–467.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- Argos, P., 1988. A sequence motif in many polymerases. *Nucleic Acids Res.* 16, 9909–9916.
- Bringloe, D.H., Pleij, C.W.A., Coutts, R.H.A., 1999. Mutation analysis of *cis*-elements in the 39- and 59-untranslated regions of satellite tobacco necrosis virus strain C RNA. *Virology* 264, 76–84.
- Canizares, M.C., Marcos, J.F., Pallas, V., 2001. Molecular variability of twenty-one geographically distinct isolates of *Carnation mottle virus* (CarMV) and phylogenetic relationship within the *Tombusviridae* family. *Arch. Virol.* 146, 2039–2051.
- Carrington, J.C., Morris, T.J., 1985. Characterization of the cell-free translation products of *Carnation mottle virus* genomic and sg RNAs. *Virology* 144, 1–10.
- Carrington, J.C., Morris, T.J., 1986. High resolution mapping of *Carnation mottle virus*-associated RNAs. *Virology* 150, 196–206.
- Carrington, J.C., Heaton, L.A., Zuidema, D., Hillman, B.L., Morris, T.J., 1989. The genome structure of turnip crinkle virus. *Virology* 170, 219–226.
- Carpenter, C.D., Simon, A.E., 1998. Analysis of sequences and predicted structures required for viral satellite RNA accumulation by *in vivo* genetic selection. *Nucleic Acids Res.* 26, 2426–2432.
- Castano, A., Hernandez, C., 2005. Complete nucleotide sequence and genome organization of *Pelargonium* line pattern virus and its relationship with the family *Tombusviridae*. *Arch. Virol.* 150, 949–965.
- Ciuffreda, P., Rubino, L., Russo, M., 1998. Molecular cloning and complete nucleotide sequence of galinsoga mosaic virus genomic RNA. *Arch. Virol.* 143, 173–180.
- Diez, J., Marcos, J.F., Pallas, V., 1999. Characterization and *in vitro* translation analysis of *pelargonium* flower break virus. *Arch. Virol.* 144, 1627–1637.
- Fabian, M.R., Na, H., Ray, D., White, K.A., 2003. 3'-terminal RNA secondary structures are important for accumulation of tomato bushy stunt virus DI RNAs. *Virology* 313, 567–580.
- Fabian, M.R., White, K.A., 2004. 5'-3' RNA-RNA interaction facilitates cap and poly(A) tail independent translation of tomato bushy stunt virus mRNA. *J. Biol. Chem.* 279, 28862–28872.
- Genoves, A., Navarro, J.A., Pallas, V., 2006. Functional analysis of the five melon necrotic spot virus genome-encoded proteins. *J. Gen. Virol.* 87, 2371–2380.
- Guilley, H., Carrington, J.C., Balazs, E., Jonard, G., Richards, K., Morris, T.J., 1985. Nucleotide sequence and genome organization of *Carnation mottle virus* RNA. *Nucleic Acids Res.* 13, 6663–6677.
- Habili, N., Symons, R.H., 1989. Evolutionary relationship between luteoviruses and other RNA plant viruses based on sequence motifs in their putative RNA polymerases and nucleic acid helicases. *Nucleic Acids Res.* 17, 9543–9555.
- Hacker, D.L., Petty, I.T.D., Wie, N., Morris, T.J., 1992. Turnip crinkle virus genes required for RNA replication and virus movement. *Virology* 186, 1–8.
- Harrell, L., Melcher, U., Atkins, J.F., 2002. Predominance of six different hexanucleotide recoding signals 3' of read-through stop codons. *Nucleic Acids Res.* 30, 2011–2017.
- Huang, M., Koh, D.C.-Y., Weng, L.-J., Chang, M.-L., Yap, Y.-K., Zhang, L., Wong, S.-M., 2000. Complete nucleotide sequence and genome organization of *hibiscus chlorotic ringspot virus*, a new member of the genus *Carmovirus*: evidence for the presence and expression of two novel open reading frames. *J. Virol.* 74, 3149–3155.
- Ishikawa, M., Meshi, T., Ohno, T., Okada, Y., 1991. Specific cessation of minus-strand RNA accumulation at an early stage of tobacco mosaic virus infection. *J. Virol.* 65, 861–868.
- Kamer, G., Argos, P., 1984. Primary structural comparison of RNA-dependent polymerase from plant, animal and bacterial viruses. *Nucleic Acids Res.* 12, 7269–7282.
- Koenig, R., Gibbs, A., 1986. Serological relationship among tombusviruses. *J. Gen. Virol.* 67, 75–82.
- Koenig, R., An, D., Burgermeister, W., 1988. The use of filter hybridization techniques for the identification, differentiation and classification of plant viruses. *J. Virol. Methods* 19, 57–68.
- Koonin, E.V., Dolja, V.V., 1993. Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit. Rev. Biochem. Mol. Biol.* 28, 375–430.
- Li, Y., Bachmann, S., Maifl, E., Commandeur, U., Breyel, E., Timpe, U., Koenig, R., 1993. Nucleotide sequence of the coat protein gene of *pelargonium* leaf curl virus and comparison of the deduced coat protein amino acid sequence with those of other tombusviruses. *Arch. Virol.* 129, 349–356.
- Liu, H.-Y., Sears, J.L., Morrison, R.H., 2003. Isolation and characterization of a carmo-like virus from *Calibrachoa* plants. *Plant Dis.* 87, 167–171.
- Lommel, S., Martelli, S.A., Rubino, L., Russo, M., 2005. *Tombusviridae*. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L. (Eds.), *Virus Taxonomy: VIIIth Report of the International Committee on Taxonomy of Virus*. Academic Press, San Diego, pp. 907–936.
- Morris, T.J., Carrington, J.C., 1988. *Carnation mottle virus* and viruses with similar properties. In: Koenig, R. (Ed.), *The Plant Viruses*, vol. 3. Plenum Press, New York, pp. 73–112.
- Marcos, J., Vilar, M., Perez-Paya, E., Pallas, V., 1999. *In vivo* detection, RNA-binding properties and characterization of the RNA-binding domain of the p7 putative movement protein from *Carnation mottle virus* (CarMV). *Virology* 255, 354–365.
- Morozov, S.Y.U., Ryabov, E.V., Leiser, R.M., Kavriev, S.K., 1995. Use of highly conserved motifs in plant virus RNA polymerases as the tags for specific detection of carmovirus-related RNA-dependent RNA polymerase genes. *Virology* 207, 312–315.
- Nam, M., Kim, S.M., Domier, L.L., Kim, Y.T., Moon, J.K., Choi, H.S., Kim, H.G., Koh, S., Moon, J.S., Lee, S.H., 2008. Soybean yellow mottle mosaic virus, complete genome. National Center for Biotechnology Information, NIH, Bethesda, MD, USA.
- Rico, P., Hernandez, C., 2004. Complete nucleotide sequence and genome organization of *Pelargonium flower break virus*. *Arch. Virol.* 149, 641–651.
- Riviere, C.J., Rochon, D.M., 1990. Nucleotide sequence and genomic organization of melon necrotic spot virus. *J. Gen. Virol.* 71, 1887–1896.
- Robertson, N.L., Cote, F., Pare, C., Leblanc, E., Bergeron, M.G., Leclerc, D., 2007. Complete nucleotide sequence of Nootka lupine vein-clearing virus. *Virus Genes* 35, 807–814.
- Russo, M., Burgyan, J., Martelli, G.P., 1994. Molecular biology of *Tombusviridae*. *Adv. Virus Res.* 44, 381–428.

- Skotnicki, M.L., Mackenzie, A.M., Torronen, M., Gibbs, A.J., 1993. The genomic sequence of cardamine chlorotic fleck carmovirus. *J. Gen. Virol.* 74, 1933–1937.
- Skuzeski, J.M., Nichols, L.M., Gesteland, R.F., 1990. Analysis of leaky viral translation codons *in vivo* by transient expression of improved β -glucuronidase vectors. *Plant Mol. Biol.* 15, 65–79.
- Song, C., Simon, A.E., 1995. Requirement of a 3'-terminal stem-loop in *in vitro* transcription by an RNA-dependent RNA polymerase. *J. Mol. Biol.* 254, 6–14.
- Stuart, G., Moffett, K., Bozarth, R.F., 2004. A whole genome perspective on the phylogeny of the plant virus family *Tombusviridae*. *Arch. Virol.* 149, 1595–1610.
- Stupina, V., Simon, A.E., 1997. Analysis *in vivo* of turnip crinkle virus satellite RNA C variants with mutations in the 3'-terminal minus-strand promoter. *Virology* 238, 470–477.
- Suzuki, S., Hase, S., Takahashi, H., Ikegami, M., 2002. The genome organization of pea stem necrosis virus and its Assignment to the genus *Carmovirus*. *Intervirology* 45, 160–163.
- Takemoto, Y., Kanehira, T., Shinohara, M., Yamashita, S., Hibi, T., 2000. The nucleotide sequence and genome organization of *Japanese iris necrotic ring virus*, a new species in the genus *Carmovirus*. *Arch. Virol.* 145, 651–657.
- Tzanetakis, I.E., Martin, R.R., 2004. Complete nucleotide sequence of a strawberry isolate of *Beet pseudo yellows virus*. *Virus Genes* 28, 239–246.
- Valverde, R.A., Nameth, S.T., Jordan, R.L., 1990. Analysis of double-stranded RNA for plant virus diagnosis. *Plant Dis.* 74, 255–258.
- Weng, Z., Xiong, Z., 1997. Genome organization and gene expression of saguaro cactus carmovirus. *J. Gen. Virol.* 78, 525–534.
- White, J.D., 2003. Calibrachoa mottle virus: Are your plants in danger? *GrowerTalks, Culture Notes*, January 2003.
- White, K.A., Skuzeski, J.M., Li, W.Z., Wei, N., Morris, T.J., 1995. Immunodetection, expression strategy and complementation of turnip crinkle virus p28 and p88 replication components. *Virology* 211, 525–534.
- You, X.J., Kim, J.W., Stuart, G.W., Bozarth, R.F., 1995. The nucleotide sequence of cowpea mottle virus and its assignment to the genus *Carmovirus*. *J. Gen. Virol.* 76, 2841–2845.
- Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415.